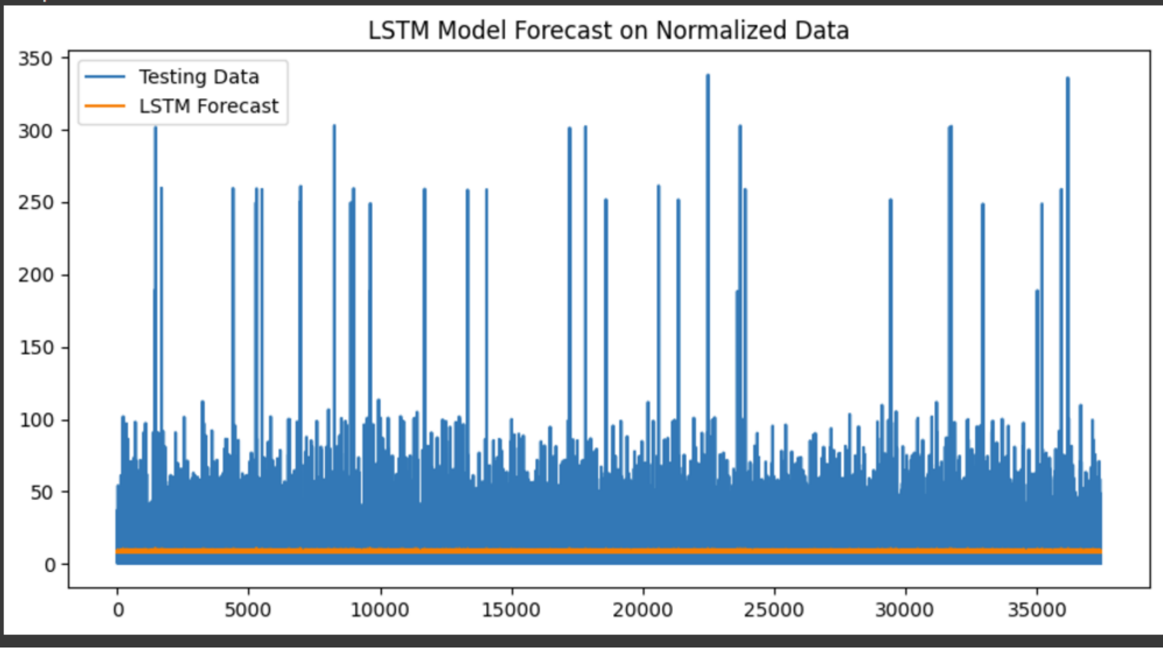
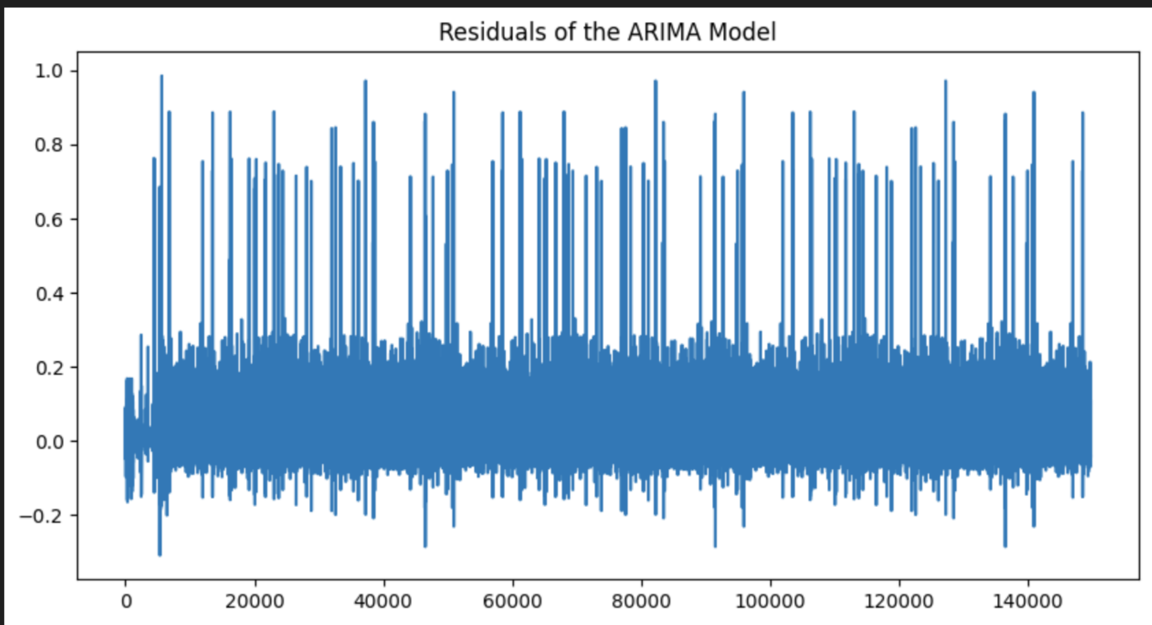
Comparison of ARIMA and LSTM Models for Price Prediction

```
Mean Absolute Error (MAE): 9.109879946949945
Mean Squared Error (MSE): 257.67459061597555
Root Mean Squared Error (RMSE): 16.05224565647983
R-Squared: -0.0007045055238354969
```




```
Mean Absolute Error (MAE): 0.028580670549926582
Mean Squared Error (MSE): 0.0022606098586956725
Root Mean Squared Error (RMSE): 0.04754587110039812
R-Squared: -0.0019666218873510832
```



Performance Metrics:

1. **ARIMA Model**:
    - Mean Absolute Error (MAE): **0.0286**
    - Mean Squared Error (MSE): **0.0023**
    - Root Mean Squared Error (RMSE): **0.0475**
    - R-Squared: **-0.002**
2. **LSTM Model**:
    - Mean Absolute Error (MAE): **9.1099**
    - Mean Squared Error (MSE): **257.6746**
    - Root Mean Squared Error (RMSE): **16.0522**
    - R-Squared: **-0.0007**

Model Comparison:

- The **ARIMA model** demonstrates significantly lower error metrics (MAE, MSE, RMSE), indicating it predicts prices much closer to the actual values compared to the LSTM model.
- Both models exhibit **negative R-Squared values**, which indicate that the models perform worse than a simple mean-based prediction. However, the ARIMA model's error is relatively smaller, making it the better choice.

Which Model Should we Use?

- **ARIMA** is the better model for this price prediction task, as it has much lower error metrics. It seems to capture the patterns in the data better than LSTM in this case.
- **LSTM** underperforms likely due to data-related issues, insufficient data size, or hyperparameter settings that do not suit the problem well.

Reasons for Poor Performance (Negative R-Squared):

1. **ARIMA Model**:
    - A negative R-Squared value means the model predictions explain less variance than simply using the mean of the target variable. This can happen if:
        - The data lacks clear trends or seasonality for ARIMA to model effectively.
        - The model parameters (p, d, q) are not optimized.
        - Preprocessing issues like non-stationary data are present.

2. **LSTM Model**:
   - The negative R-Squared for LSTM suggests that the model is failing to generalize. Possible reasons include:
     - Insufficient data: LSTMs require a large dataset to perform well.
     - Poor preprocessing: Lack of normalization or incorrect sequence length for inputs.
     - Overfitting: The model may memorize the training data without learning patterns in the test data.

How Can I Improve Both Models?

Improvements for ARIMA:

1. **Parameter Optimization**: Use a grid search to find the best values for (p, d, q) or leverage the Akaike Information Criterion (AIC) for model selection.
2. **Incorporate Exogenous Variables**: If additional features like seasonal indicators or external factors (e.g., holidays) are available, include them in the ARIMA model.
3. **Resampling**: If the data has irregular time intervals, resample it to ensure consistency.

Improvements for LSTM:

1. **Data Preprocessing**:
   - Experiment with different sequence lengths for input data to capture the right temporal dependencies.
2. **Architecture Tuning**:
   - Adjust the number of layers, neurons, and activation functions.
   - Add dropout layers to prevent overfitting.
3. **Training Process**:
   - Use techniques like early stopping to avoid overtraining the model.
   - Experiment with learning rates, optimizers (e.g., Adam), and batch sizes.
4. **More Data**:
   - If possible, increase the dataset size, as LSTM models typically perform better with larger datasets.

Final Recommendation

For this task, I recommend using the **ARIMA model** as it shows better performance based on the metrics. we should focus on improving the ARIMA model by ensuring data preprocessing is thorough and optimizing the parameters. While LSTM has the potential for better performance in capturing non-linear relationships, it requires a larger and more complex dataset to outperform ARIMA.

Both models have room for improvement, but the choice ultimately depends on the characteristics of the data. If the data contains clear trends and seasonality, ARIMA will perform well. If the data is non-linear and complex, an optimized LSTM (with sufficient data) could be revisited.